



www.sciencemag.org/cgi/content/full/science.aap7911/DC1

Supplementary Materials for

Oklahoma's induced seismicity strongly linked to wastewater injection depth

Thea Hincks, Willy Aspinall, Roger Cooke, Thomas Gernon*

*Corresponding author. Email: thomas.gernon@noc.soton.ac.uk

Published 1 February 2018 on *Science* First Release
DOI: 10.1126/science.aap7911

This PDF file includes:

Materials and Methods
Figs. S1 to S18
Tables S1 to S7
References

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/cgi/content/full/science.aap7911/DC1)

Data File S1: [aap7911_s1.csv](#)

Materials & Methods

Bayesian Network development

Bayesian Network (BN) analysis was performed using the UNINET COM library (33) through R (34). UNINET is designed for high dimensional dependence modeling and multivariate ordinal data mining using BNs. UNINET supports non-parametric BNs with continuous, discrete, and functional nodes, and is unique in that it represents conditional dependencies by conditional rank correlation, using the joint normal copula (46, 47). This approach makes very large and complex networks computationally tractable. Here we use UNINET in ‘data mining’ mode to estimate the best-fit network from the input data. The BN parameters are estimated entirely from data, with no prior assumptions other than specifying the node hierarchy. This dictates the direction but not the magnitude of influence between nodes (Table S2), which are treated as continuous variables.

Data handling

Monthly injection records for Underground Injection Control (UIC) wells in Oklahoma (summarized in Fig. 1b; Figs. S1 and S2; Table S1) were obtained from the Oklahoma Corporation Commission, OCC (35). The primary BN analysis presented here uses records from January 2011 to December 2015 (the most recently available at the time of writing). During the review process, data for 2016 became available and were used to test both BN forecasting and model updating.

The majority of active (i.e. reported volume >0 bbl) UIC wells in Oklahoma lie within the ‘Oklahoma induced seismicity zone’ (Fig. 1A and Table S1) (5). Our analysis is restricted to wells within this zone.

Wells are identified by a unique American Petroleum Institute number (API). There are some missing or erroneous entries in the records, and we have made every attempt to rectify these. For

example, there are instances where records with the same API (i.e. pertaining to the same well) have missing or differing coordinates, some with zero or unfeasibly small depth values, and a number of duplicate entries (which have been removed). We also note that some APIs appear to have been modified to accommodate new (neighboring) wells, with extra digits appended. These records, together with any missing or incorrect coordinates, have been corrected by cross checking with records from earlier years and the OCC database (35). Any records that could not be properly resolved have been removed. In some cases, it is unclear if injection records are incomplete or if injection was suspended for a period—for example, a well with records for 2011–2013 and 2015, but no data for 2014. In such cases, we have assumed that records are accurate as published. Should OCC issue corrections or additional data the BN can be updated.

Types of class II UIC wells

Our analysis includes class II UIC wells involved in saltwater disposal (SWD) and Enhanced Oil Recovery (EOR). We developed a test BN to explore the impact of different well types using all the updated and newly released datasets (2011–16) (35), including an additional node for well type (1 = SWD well, 0 = EOR well). The conditional rank correlation for well type and moment release (i.e. the correlation between well type and moment given all the other operational and spatial variables) was very low (-0.02), suggesting that well type adds very little information given our present network configuration. Therefore, at present we cannot resolve what is due to the actual well operation and what is related to local geologic conditions (faults, permeability etc.), but it is highly probable that the geospatial correction (discussed below) adequately accounts for such latent factors. One important observation is that individual SWD injection wells in the studied region inject on average ~ 1.8 times more than enhanced oil recovery wells. This is noteworthy given that active SWD wells are also > 1.5 times more likely to be associated with an earthquake than active EOR wells (4). This may be simply due

to the fact that SWD wells are injecting higher volumes, but we stress that depth and location are important confounding factors.

Description of variables used in the BN

The reader is referred to Table S2 for a complete list of variables used in the BN analysis.

Well depth

Our analysis uses the reported total well depth (node *Depth*, m). It must be noted that this does not necessarily correspond to depth of injection—wells can be plugged back, and depth of injection may change over time. However, total depth is the only measure that is consistently provided for all wells in the OCC database. For a very small number of localities (insufficient for our analysis), plug back or actual injection depth is reported.

We restrict our analysis to wells ≥ 1 km deep for two reasons—firstly to exclude erroneous data and secondly to improve forecasting for deeper and higher volume wells associated with higher seismicity.

In the OCC data files there are numerous records with unfeasibly low depths (zero or tens of meters). As disposal typically targets deep geologic formations, these do not appear to be legitimate. From a spot check of several <100 ft wells using the online OCC database, these records appear to be missing data (no depth recorded) or missing a decimal place. Wells with a reported depth of 999 ft also appear to correspond to missing data. Without manual checking, the minimum cutoff to exclude missing or erroneous depths would therefore be 1000 ft or 305 meters.

Deeper (and higher volume) wells are understood to be associated with a greater risk of induced seismicity. We replicated the BN using just records for wells <1 km deep, and observed a very weak correlation between well depth and \log_{10} annual moment release (empirical rank correlation of 0.006), and monthly volume and \log_{10} annual moment release (-0.002), supporting

the argument that these wells have a considerably lower impact on seismicity.

Basement depth

Basement depth was estimated by performing a surface interpolation (regularized spline with tension, using the GRASS GIS function `v.surf.rst`) using records of 1232 wells drilled into the crystalline basement (39). The resulting raster map of approximate basement depth (Fig. S5) was sampled at each UIC well location to calculate well depth relative to basement (node *Depth_rel_InterpB*, m) and the absolute distance from the bottom of the well to the basement interface (node *distBasement*, m, this is always positive). The accuracy of this interpolation will vary spatially due to variable data coverage across the state. However, the strong correlation observed between moment and these two measures (particularly distance to basement, see Table S5) suggests that the basement estimate is reasonably accurate in the area of interest, and that there is valuable information in these parameters.

Volume

For each well report, records of monthly injected volume (node *Volume*, bbl/month) were used to calculate total injection over the past year (node *Vsum_1y*, bbl/year), and cumulative volume for every month of operation (node *CumVolCalc*, total bbls injected up to the given report date).

We then filtered the records to omit cases with injection $< 10,000$ bbl/m. Our analysis focused on wells that are or could be subject to regulation in the future, and we removed cases when the injection rate (or depth, see above) means that they are not, or are very unlikely to be, related to earthquake activity. This is evidenced by low correlations for volume and \log_{10} moment computed for records $< 10,000$ bbl/m only (an empirical rank correlation of 0.001). For these very low volume wells, local seismicity is likely to be influenced to a much greater extent by the contribution of neighboring wells (particularly high-volume, deeper wells). Eliminating cases that are very weakly correlated with seismic moment release (and as a result, many

low/zero moment cases) improves forecasting performance for the higher volume/deeper wells.

We have only omitted *individual* cases (monthly reports) where injection is $<10,000$ bbl, so a well is only completely excluded from the analysis if it never injected over 10,000 bbl/month in the study period. Injection $<10,000$ bbl/month still contributes to annual and cumulative volumes. The regulatory threshold for Oklahoma is 15,000 bbl/day ($\sim 450,000$ bbl/month) (25) so our cutoff still includes injection at the lower end of the scale.

Earthquake activity close to any given well could be due in part (or whole) to injection from nearby wells. However, evaluating the potential contribution of neighboring wells is difficult, as injection depth and volume appear to be jointly important. We ran a number of tests using a range of variables to represent external inputs (including the number of active neighboring wells, and total volume injected within different radii and depths). All the nodes tested provided very little additional information (low empirical and conditional rank correlations) and did not improve forecast performance. However, we note that it is very difficult to deal adequately with the effect of depth (and basement depth) when evaluating the effect of the total injection within a given area. This would necessarily involve some heavy discretization and approximation.

The geospatial correction node will to some extent capture the impact of well density, and spatial variation in total injection and depth.

Seismic moment release

Earthquake depth and magnitude data were obtained from the Advanced National Seismic Systems' (ANSS) Comprehensive Earthquake Catalog, ComCat (36). We investigated the minimum magnitude of completeness (M_c) for earthquake data for the Oklahoma induced seismicity zone from 2011–2016, and a more recent subset of data (2015–2016), using the approach of Wiemer and Wyss (48). We found that the M_c for 2011–2016 is 2.50 ± 0.025 (1 s.d.) and for 2015–2016 is 2.52 ± 0.039 (1 s.d.) (Fig. S4).

Catalog incompleteness means moment is censored, and the distribution for total annual

moment release is not smooth at small values. To handle this, we applied a minimum threshold of $1e13$ Nm (assigning this value to all cases with annual moment $<1e13$ Nm), approximately equivalent to one M_w 2.6 event. This truncated distribution forestalls any effect of missing small events below the threshold of completeness. We also stress that our models used total seismic moment release, which is predominantly determined by larger events, thus catalog completeness becomes a marginal consideration and has very little effect on the results.

A modified version of rcomcat (49) was used to search the database for earthquakes ($M_w \geq 2$ and ≤ 10 km deep) within 20 km of each disposal well. For each individual record we then calculate *lmom_1y*, the total moment released in a radius of 20 km in the following year. Other studies encompassing the CEUS use a 15 km ‘distance of association’ (4), which sums the 5 km radius traditionally regarded as induced (50) and the estimated 10 km spatial uncertainty in the earthquake epicenter location (51). Our limit is intended to capture earthquake swarms (e.g. the Jones swarm) that are known to occur within 20 km of high-rate disposal wells (3). McGarr and Barbour (37) concluded that for the 2016 Fairview, Pawnee, and Cushing earthquakes, 10 km was the minimum distance required to capture seismic moment release and the injection volume. We performed additional tests focusing on a smaller region (around Prague) to evaluate the rank correlation coefficients for injected volume (monthly, annual and cumulative) and total annual moment release using varying radii from 5–35 km. These showed the strongest correlations at distances of 15–20 km—further supporting this choice of radius.

Geospatial correction

The geospatial correction is one of a class of statistical modeling constructs known as “latent variables”. Such a variable is not directly observable or observed, but can be inferred and characterized by analyzing the variance and covariance of related effect indicators. Latent variables are often used in, e.g., in Hidden Markov processes, mixed effect/multi-level, or regression equation modeling; a latent variable also has a structural role in causality (52). The geospatial

correction factor likely accounts for multiple physical processes or rock properties that are not determinable at the scale of individual wells, including: pore pressure differential; fault friction; strata permeability, porosity, hydraulic diffusivity; shear modulus, Poisson's ratio; poroelastic and other rheological properties.

The geospatial correction is obtained by generating a kernel density map of the forecast error produced by the first stage (non-spatial, 7-node) BN. This network is learned using a subset of the training data, then used to estimate seismic moment for the remainder of the training data. The resulting forecast error (i.e. the difference between the observed and forecast seismic moment) is used to generate the kernel density, giving a measure of regional under- or over-forecasting. We then take point samples from this geospatial correction map (Fig. S8) to estimate the appropriate correction variable for each individual well, using this sampled value as a 'new' node in the second stage, 8-node geospatial BN (Fig. 2). Fig. S8 shows the geospatial correction map for the unsaturated BN (using the Test 2 training dataset). This procedure was repeated to generate equivalent corrections for both the saturated BN and linear regression model. We note that all three models give a very similar spatial distribution, and that a similarly large improvement was seen in the regression model after incorporation of this latent variable. Updating the geospatial correction (see section on forecasting using 2016 data, at the end of this supplement) also appears to partially accommodate temporal changes in the system.

Analysis of the BN: Model performance, influence of geospatial correction and node hierarchy

Tests were performed using the saturated BN (structure shown in Fig. 2A) to explore the strength of correlation between variables, and the effect of the geospatial correction on model performance. Test 2 data were used for learning, and a random sample of 16,000 cases from the learning set used to test performance (see Table S3, in-sample test).

Fig. S9 shows a scatter plot of forecast versus observed annual moment release for the saturated network and regression models with no geospatial node. It can be seen from the CDF and RMSE values (Fig. S9B and table inset in S9A) that the BN performs better than the regression, but that neither model is particularly good, particularly for cases with high moment release. Fig. S9C shows that the conditional standard deviation of moment release (the uncertainty in the forecast) is strongly dependent on expected moment release.

Both the BN and regression model show a marked improvement with inclusion of the geospatial correction, and this can be seen in the scatter plots and RMSE (compare Fig. S10 A and Fig. S9A). Fig. S10C shows that the conditional standard deviation of moment release (the uncertainty in the forecast) still varies strongly with expected moment release but is slightly lower than for the network without the geospatial correction.

Correlations and inferred effects of depth and volume on seismic moment release for individual wells

Table S5 presents the unconditional rank correlations calculated using UNINET for (i) the empirical data; (ii) the saturated BN; (iii) the unsaturated BN; and (iv) the linear regression correlation coefficients. The empirical rank correlation coefficients show that after the geospatial correction, the distance from the bottom of the well to the estimated basement interface (*distBasement*) is the variable most highly (negatively) correlated with seismic moment release (rank correlation coefficient of -0.34 for the empirical data and saturated BN, -0.32 for the unsaturated BN). This suggests there are more, and/or larger magnitude earthquakes associated with wells that terminate close to the sedimentary cover–basement interface. This can also be seen in the empirical probability distributions for *depth relative to basement* and *distance to basement* (Fig. S7), where we compare the probability density for cases with no or low seismicity (\log_{10} annual moment ≤ 13 , shaded blue) with relatively high seismicity cases (\log_{10} annual moment ≥ 15 , shaded pink). The distributions show multiple peaks and troughs as

the data are non-Gaussian and granular, but the overall shift is clear. The probability densities for both distance to basement and depth relative to basement for high seismicity (\log_{10} annual moment ≥ 15) both peak close to zero (corresponding to wells approximately at the basement interface) and are much narrower than the equivalent distribution for low/no seismicity.

Cumulative volume has an empirical rank correlation of +0.26—suggesting a higher degree of influence on moment release than either annual injection (empirical rank correlation of +0.20) or monthly volume (empirical rank correlation of +0.18). This could be indicative of a relatively long time lag associated with increased fluid pressure and stress accumulation on faults, at least in some areas. Again, all the probability distributions for volume (monthly, yearly and cumulative, Fig. S7) show that higher volumes are associated with higher seismic moment release, with cumulative volume showing the most distinct shift.

Influence of node hierarchy on *conditional rank correlation*

The empirical and unconditional BN rank correlations (Table S5) show distance to basement to be the single well parameter with the strongest influence on moment release. However, these unconditional rank correlations do not account for confounding effects of other variables and are therefore a relatively crude indicator of importance.

Partial (or conditional) rank correlations given all other covariates take account of confounders, and therefore provide a measure of the additional information provided by that node alone. For example, the partial correlation of (x, y) given u, v, w takes account of confounding by u, v, w ; i.e. the influence between x, y that is not caused by u, v, w . The partial correlation of (x, y) given v, w therefore takes account of confounding by v, w , but u is still a potential confounder. The absolute rank correlation of (x, y) does not account for the confounding effect of any other variables.

Table S4 gives partial rank correlations for two alternative node hierarchies, showing that

node order has little effect on the strength of correlation for the distance to basement node, and it remains the highest correlated variable. The partial correlation for distance to basement and $lmom_1y$ given **all** other depth and volume variables is -0.28 (compare with the absolute rank correlation of -0.34, Table S5). The effect of confounders (at least in terms of the variables in our network) on the influence between distance to basement and $lmom_1y$ is therefore relatively small.

The non-parametric BN representation captures the joint distribution of a child node and its parents in terms of a set of conditional rank correlations or equivalently (in the presence of the Gaussian copula) partial rank correlations. If node variable W has parents X, Y, Z , then with the margins of W, X, Y, Z , and conditional on all parents of X, Y, Z , the joint distribution of W, X, Y, Z is completely specified by specifying the partial rank correlations:

$$r(W, X), r(W, Y | X), r(W, Z | X, Y).$$

These partials are algebraically independent: they may be chosen independently in the interval $(-1, 1)$.

The same distribution could be specified by giving:

1. $r(W, Y), r(W, X | Y, X), r(W, Z | X, Y)$ or
2. $r(W, Z), r(W, Y | Z), r(W, X | Z, Y)$, etc.

In the latter sense, the ordering in the hierarchy is unimportant. However, it is important if we are interested in specific variables and their potential confounders. Then, in respect of the relation between W and X , say, we should choose option (2) to get the relation between W and X which is not confounded by the influence of Z and Y . If we are interested in the relation W, Y , then we should choose (1). However, this issue is purely cosmetic as (1) and (2) specify exactly the same joint distribution. The difference is that the unconfounded relation of (W, Y) can be ‘seen’ in the first, and the unconfounded relation of (W, Z) in the second.

Training and testing datasets

We test the models with two alternative subsets of the data to evaluate performance, illustrate how the BN could be used operationally (Fig. S11), and also to investigate how the model responds to the changing trend of injection seen in 2015. The number of cases used for learning and testing or forecasting in each instance (including the number of large moment cases) are given in Table S3.

Test 1 demonstrates how the BN can be used to forecast ‘ahead’, learning with the first phase of observations from January 2011–June 2015, and forecasting for the period July–December 2015. This phase is interesting, as from 2011–2014 annual injection steadily increased, but began to decline at the start of 2015 (Fig. S1). The reduction in injection can be attributed to the introduction of more stringent regulations, and falling (US) oil production volumes due to low oil prices.

Test 2 uses 90% of the full dataset (randomly sampled, from January 2011–December 2015) for learning, and the remaining 10% for forecast verification. Predicted moment is then compared with observations (for both BN and regression models) to assess how well the models perform.

Performance of alternative network configurations

We present results for two different network configurations: (i) a saturated network (Fig. 2A), where the network is ‘complete’, i.e. there are links between every pair of nodes (no enforced conditional independence); and (ii) an unsaturated network (Fig. 2B), which allows conditional independence between certain parameters.

The saturated BN gave the most complete representation of all the observed data under stationary conditions, and therefore initially appears to better represent the dataset as a whole. Using the Test 2 dataset (where the test data are effectively drawn from the same statistical

population as the learning data) the saturated BN gave the lowest RMSE (Table S6), and the saturated BN rank correlation coefficients for *lmom_1y* were closest to the empirical rank correlations (Fig. 2C).

In an operational or regulatory setting it is necessary to make projections for the future, under potentially quite different conditions (e.g. assessing the potential impact of new wells, or restricting injection rates). Under certain changing conditions, the saturated BN can produce ‘unphysical’ results, and it is necessary to remove links between the volume nodes to enforce physically realistic constraints on dependencies between the variables. Test 1 (and our additional updated forecast using provisional injection data for 2016, presented at the end of this supplement) are examples of forecasting ahead, where the test dataset is statistically quite different to the learning set — in this case, both as a result of the changing injection regime and a proportional increase in the number of cases with higher moment release (see Table S3).

The trend of increasing injection from 2011 to end 2014 results in strong (and dominant) correlations between cumulative, annual and monthly injection volumes in the saturated network (Table S5). When injection rates fell at the end of 2015, these correlations changed and as a result the saturated network performs significantly less well than both the unsaturated BN and regression models. The unsaturated BN, which enforces conditional independence between cumulative, annual and monthly injection, gives the best performance, with the lowest RMSE, **MAE and MAD** (Table S6). This can also be seen in the CDF plots for the unsaturated network—the BN median estimate very closely follows the observations (Fig. S12C).

Removing the links between the volume nodes in the unsaturated BN is therefore a way to (partially) compensate for the injection regime changes of 2015 (see Fig. S1). As such this is a more reliable model for ongoing forecasting, future regulatory assessment or quantifying the impact of new wells.

Quantitative assessment of depth and volume regulation

(i) Reducing injection depth by 1000 m

This is a simple assessment of a fixed change in depth, using the unsaturated BN with no geospatial node (i.e. a ‘generic’ case, focusing on the joint effects of well and basement depth and volume, excluding the confounding effects of location-specific geologic features). This analysis uses a subset of 10% of the well records, randomly selected to provide a representative sample of real operating depths and injection rates across the Oklahoma induced seismicity zone. For each record, well depth was reduced by 1000 m and the corresponding depth relative to basement and distance to basement re-calculated. The BN was then used to generate mean estimates of moment release using both the original and reduced depths (Fig. 4A and Fig. S14).

(ii) Capping injection volume

Here we evaluate the impact of a daily (or monthly) cap on injection volume, using a regulatory limit of 15,000 bbl/day (~450,000 bbl/month). Records used in this example are cases of high injection only (where monthly injection was reported as exceeding 450,000 bbl/month—a total of 2,015 records). We are therefore only comparing moment estimates for those months where the cap would be in effect. Again, this is a ‘generic’ example, using the unsaturated BN with no geospatial node.

(iii) Statewide impact of limiting injection depth relative to the basement

Here we demonstrate how the BN can be applied to quantify the expected change in annual seismic moment release if well depth is restricted to either (i) 200 m or (ii) 500 m above the basement (where the basement interface is deep enough for this to be practically enforced). Here we apply the unsaturated BN with the geospatial node to model observed spatial variation across the Oklahoma induced seismicity zone. This provides an indicative estimate of the total

annual change in moment in the zone.

Using a single month of well records (December 2015), the BN is used to predict total annual moment release within 20 km of each well, injecting at its current depth. Wells with a depth relative to basement of less than 200 m are then moved upwards to 200 m above the basement. To ensure a sensible range of injection depths, depth is only modified if the basement is 700 m or lower—this results in a minimum well depth of 500 m and means a small number of wells remain within 200 m of the basement. For the 500 m limit, any well with a depth relative to basement of less than 500 m is set to 500 m, if the basement is 1000 m or lower. This gives an indicative reduction in total annual moment release for the Oklahoma induced seismicity zone of a factor of ~ 2.8 for the 500 m limit, and a factor of ~ 1.4 for the 200 m limit.

Comparison of Gaussian and optimal copulae using AIC

UNINET uses joint normal copulae to represent conditional dependencies between variables, as conditioning can be performed analytically (this is not possible for many copula families). The moment node, $lmom_1y$ is censored (due to catalog incompleteness below M_w 2.5; Fig. S4) and very skewed, which means that the joint normal copula can only approximate the original data. To investigate how significant this limitation is, the R package VineCopula (53) was used to calculate AIC (Akaike Information Criterion) for alternative models using (i) Gaussian and (ii) optimal copulae. The lower the value of AIC the better the model. For a given set of observations, the function RVineStructureSelect() searches for the optimal structure and copulae (from the full set of bivariate copula families provided in the R VineCopula package (53)), using AIC as the scoring metric.

Using the training dataset (90% of the full set of observations, randomly sampled, Fig. S11) and 8 variables (including the geospatial correction) the Gaussian model gives an AIC of -1975573 , and the model using optimal copulae a slightly lower AIC of -2067686 . For

comparison, AIC for the regression model (again with 8 variables and the same input dataset) is significantly higher at -71063.5 . We note that the Gaussian copula is the only one that enables rapid conditionalization on *arbitrary* sets of conditioning variables. Given the relatively small improvement in AIC, non-Gaussian alternatives did not seem worth pursuing here. Cooke et al. (54) provide a detailed illustration of vine regression, using both Gaussian and optimal vines.

Testing the BN using new data

A distinct advantage of the BN as an operational tool is the ability to update the model, and produce revised forecasts as new data become available. During the review process for this contribution, injection data for 2016 were released by the OCC (35). This provided an opportunity to test our model using a new dataset. We note that these test data for 2016 are preliminary, and require further checks to ensure data reliability and consistency. For example, around 1000 wells have revised API numbers (to accommodate new wells). We have (conservatively) eliminated any cases where individual well API and location could not be uniquely resolved and traced back over the full reporting period.

Here we present two examples of forecasting using injection records from January to September 2016 (as the BN estimates seismic moment release in the year ahead, i.e. up to end September 2017). The first example uses the original model and geospatial correction derived from 2011–2015 records (Fig. S16). Updated calculations of the empirical and BN unconditional rank correlations for moment release using the updated dataset show relatively small changes in the correlations with the addition of the 2016 data (see Table S7, and Table S5 showing the original correlations for the 2011–15 data), giving confidence in our original findings.

Secondly we demonstrate how temporal changes in the system can be partially accommodated by updating the geospatial correction and using a moving window for learning (Fig. S17), improving forecast performance. This is a more tractable alternative to a fully dynamic BN.

The forecasting window

The BN uses a ‘rolling’ window for forecasting, enabling a year ahead forecast to be generated at any point in time. This also accommodates (to some extent) uncertainty/variation in lags between injection and induced seismicity.

With the time window adopted for forecasting, there is necessarily some overlap of monthly forecasts. Thus, when testing ability to forecast ahead, there is also some overlap in the earthquake records used to calculate total moment in the training and testing datasets. To demonstrate that this overlap does not artificially increase forecast performance, we calculated the variation in forecast error over time for 2016 (Fig. S18). This example used injection data from January 2014 to December 2015 to learn the network (see Fig. S17), and a geospatial correction calculated from the previous forecast for 2015 (Fig. S12), reflecting an update to the original network and geospatial variable. Boxplots (Fig. S18A-B) show monthly forecast residuals (forecast – observed \log_{10} moment) for both the BN and the regression model for January to September 2016. The test set comprised a total of 11,000 individual records (~ 1250 /month) forecasting moment release for the year ahead, so including activity to end September 2017. This test shows that the various measures of forecast error (RMSE, MAE and MAD, Fig. S18C-D) remain relatively constant.

The length of overlap between the time windows used to sum moment release in the learning set (January 2014 to December 2016) and the 2016 testing set (January 2016 to September 2017) decreases with time. If this were a major contributory factor to model forecast ‘skill’, the errors could be expected to increase with time: this is not observed.

Supplementary Figures

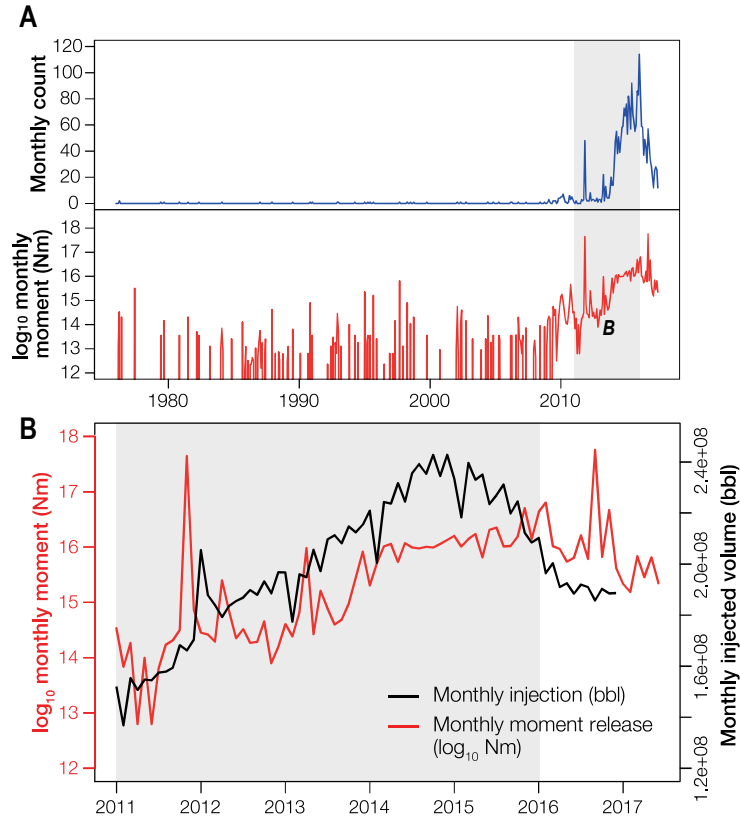



Figure S1: Seismic and UIC injection records for Oklahoma | (A) Monthly number of $M_w \geq 3$ earthquakes and total moment release (\log_{10} Nm) for $M_w \geq 2$ earthquakes for 1976–2017 (36). Note that since mid-2015, count has dropped—but total moment remains high—implying fewer but large earthquakes. (B) Total monthly injection volume (black) across the State of Oklahoma for 2011 to end 2016 (35), and monthly total seismic moment release (red, $M_w \geq 2$) for Jan 2011 to Sep 2017 (36). The shaded area highlights the time period of injection from Jan 2011 to Dec 2015 used in the primary BN analysis. Preliminary injection records from Jan to Sep 2016 were used to perform additional tests of forecasting ability and to demonstrate model updating (Figs. S16 and S17). Annual injection volumes are provided in Table S1.

	Number: 1	Author: vrobin03	Subject: Highlight	Date: 2/6/2020 2:54:13 PM -07'00'
Figure S1: Seismic and UIC injection records for Oklahoma				

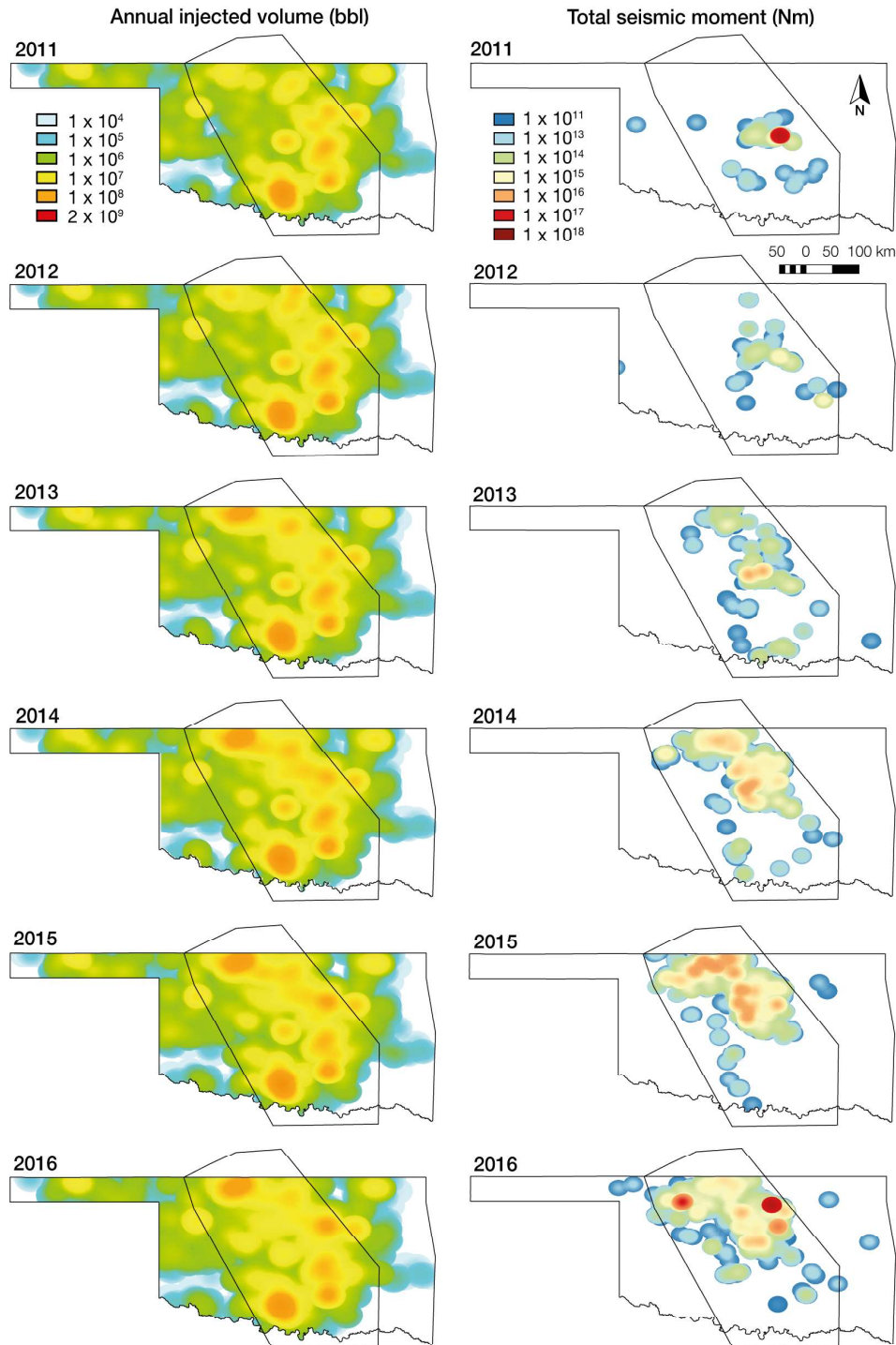


Figure S2: **Annual wastewater injection and seismicity in Oklahoma** | Maps show the total annual injected volume of wastewater (bbl) from 2011–2016 (35) using a 30 km radius Gaussian kernel, and total seismic moment release (Nm) from 2011–2016 (36), using a 20 km radius Gaussian kernel.

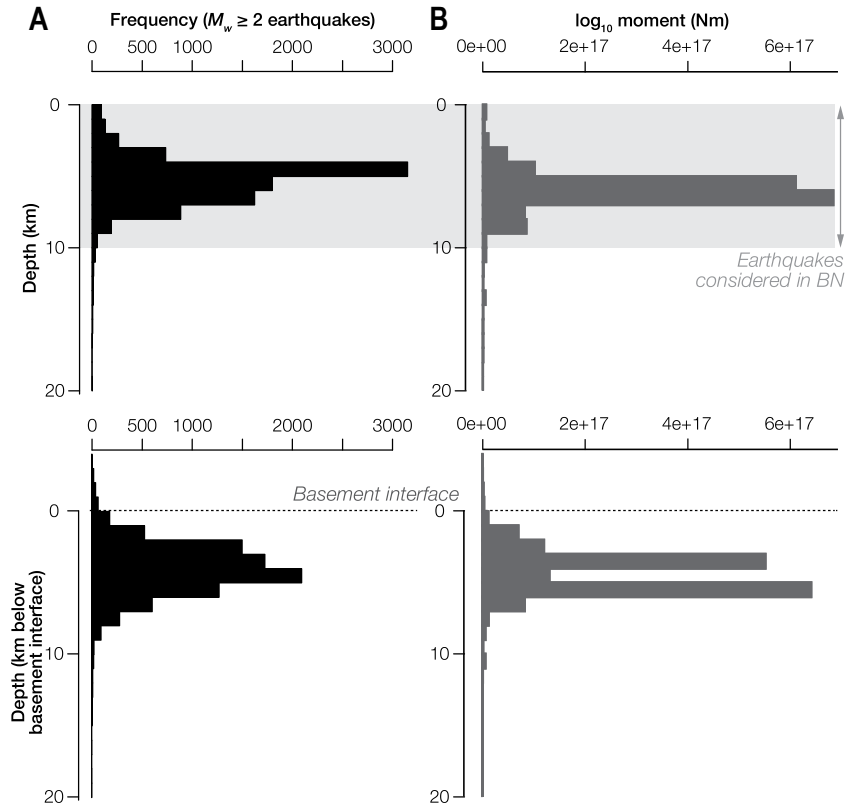


Figure S3: **Depths of Oklahoma earthquakes from 1976–2017** | Depth histograms of earthquakes ($M_w \geq 2$) from January 1976 to June 2017 (36), showing (A) frequency (count) and (B) total moment release. For the lower two plots, depth is calculated relative to the (estimated) basement interface (see Fig. S5).

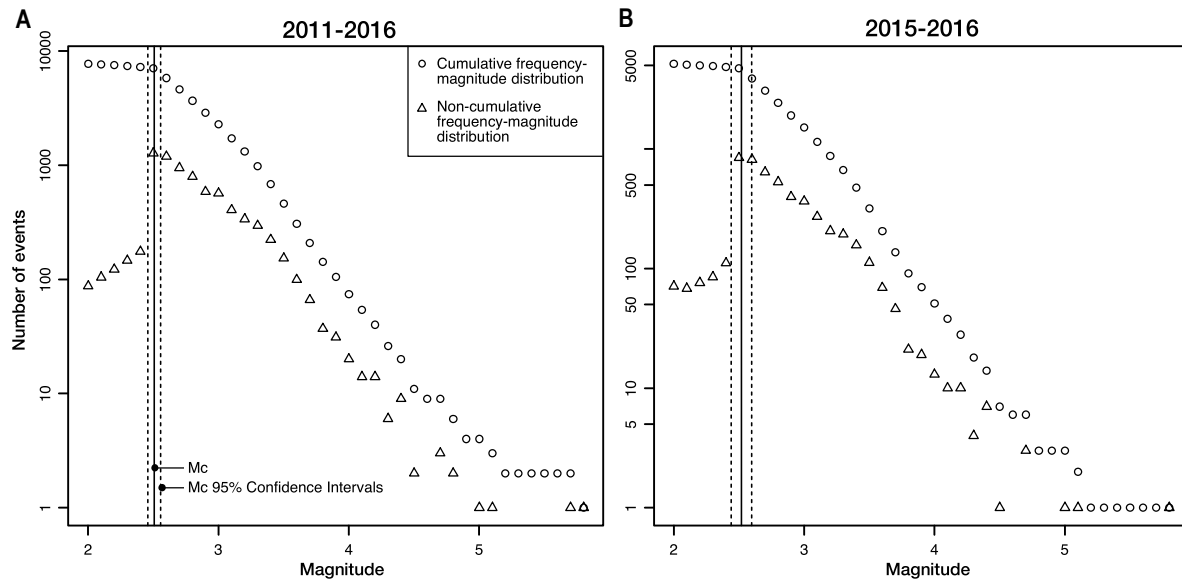


Figure S4: **Minimum magnitude of completeness (M_c) in the Oklahoma earthquake catalog** | M_c (denoted by the solid vertical line) was calculated using the method of Wiemer and Wyss (48) for earthquake data from the Oklahoma induced seismicity zone from (A) 2011–2016, and (B) a smaller subset from 2015–2016. Cumulative and non-cumulative frequency-magnitude distributions are shown. The M_c 95% confidence intervals are also plotted as vertical dashed lines (labeled).

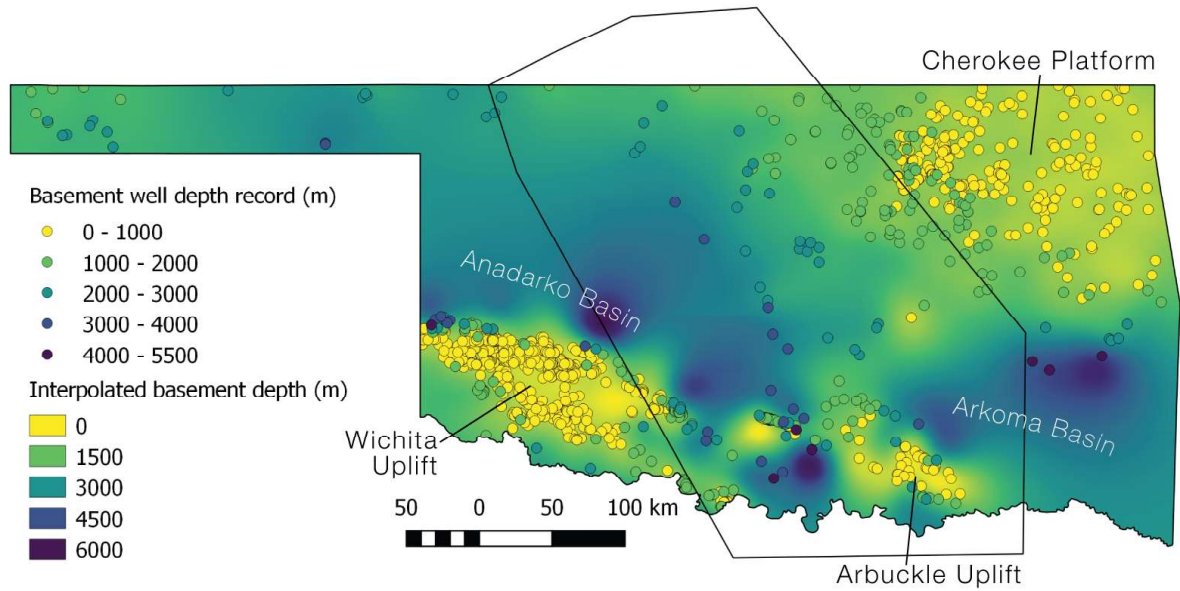


Figure S5: Depths to the basement interface across the State of Oklahoma | This compilation includes 1232 depth records of wells drilled to crystalline basement (igneous and metamorphic lithologies) from Oklahoma Geological Survey (39). Basement depths were estimated using a surface interpolation algorithm (regularized spline with tension) in QGIS (38), and are relative to surface elevation. A total of 335 depth records are within the ‘Oklahoma induced seismicity zone’ (5), shown as a black polygon.

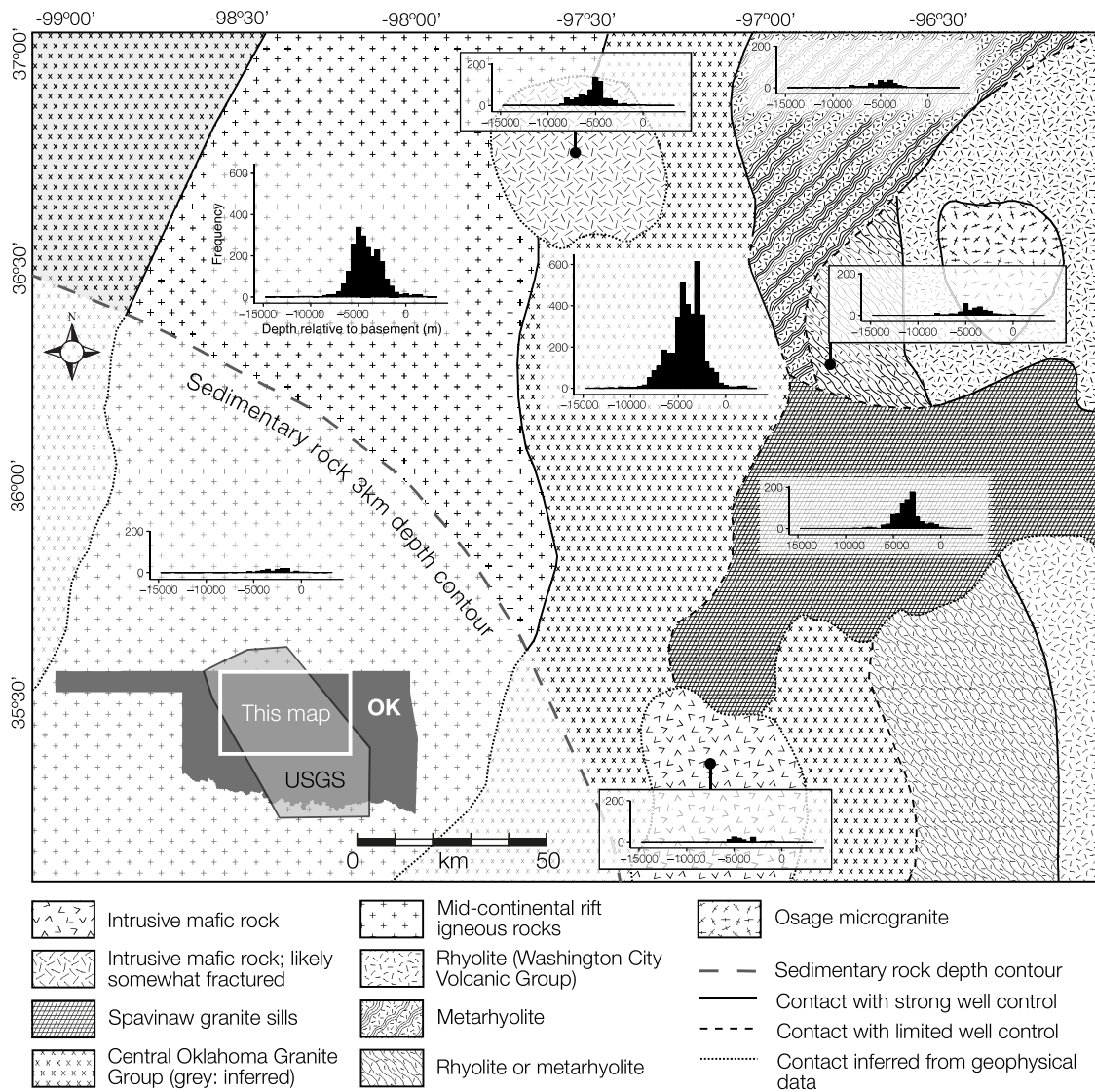


Figure S6: Oklahoma crystalline basement geology and depths of earthquakes relative to the basement interface | This region lies in the northern part of the Oklahoma induced seismicity zone (5) (see inset). Basement geology is modified after Shah & Keller (9) and is based on a combination of gravity, magnetic and drill core data. Histograms show the frequency of earthquakes ($M_w > 2$) for the period 1976 to June 2017, and are calculated relative to the (estimated) basement interface (see Fig. S5), with negative numbers referring to depths below the basement interface. Note that each histogram corresponds to a discrete basement lithology, and assumes that geologic boundaries are vertical. It can be seen that seismicity is largely confined to the Central Oklahoma Granite Group and mid-continent rift igneous rocks (9), where it is largely focused between 0–10 km below the basement interface (Fig. S3).

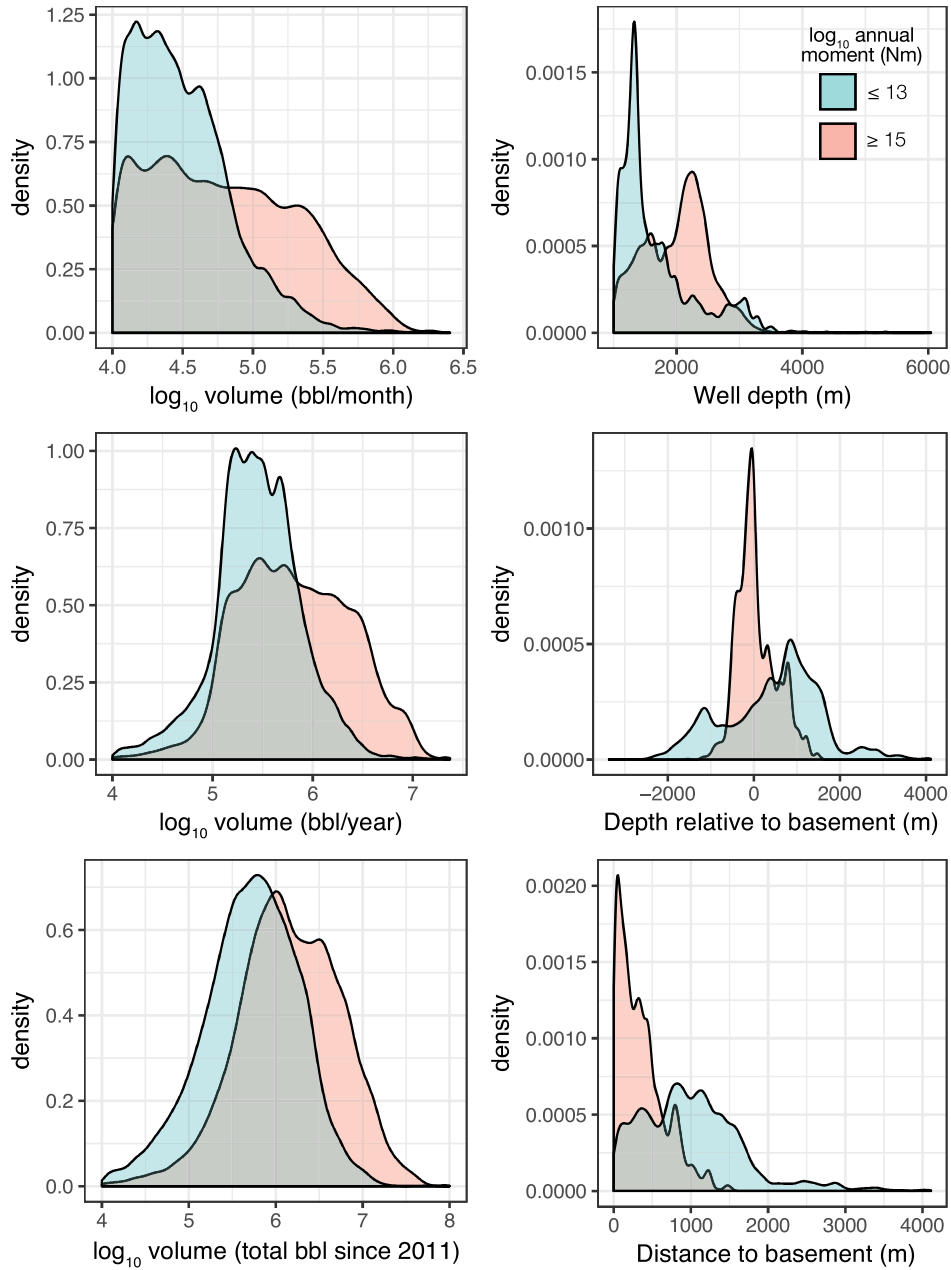


Figure S7: Probability distributions for BN input variables | Probability distributions for each of the BN input variables for January 2011 to December 2015 identifying (i) cases with low or no seismic activity (\log_{10} moment in 1 year ≤ 13 ; density shaded blue) and (ii) cases with relatively high seismic activity (\log_{10} moment in 1 year ≥ 15 ; density curve shaded pink). This shows that wells drilled closer to the basement and injecting greater volumes (over monthly, yearly and longer timescales) are associated with higher seismic moment release. Note: ‘Cumulative volume’ is only summed over the period for which injection data were available at the time of writing (2011 to end 2015) so here we present injection over a maximum of 5 years.

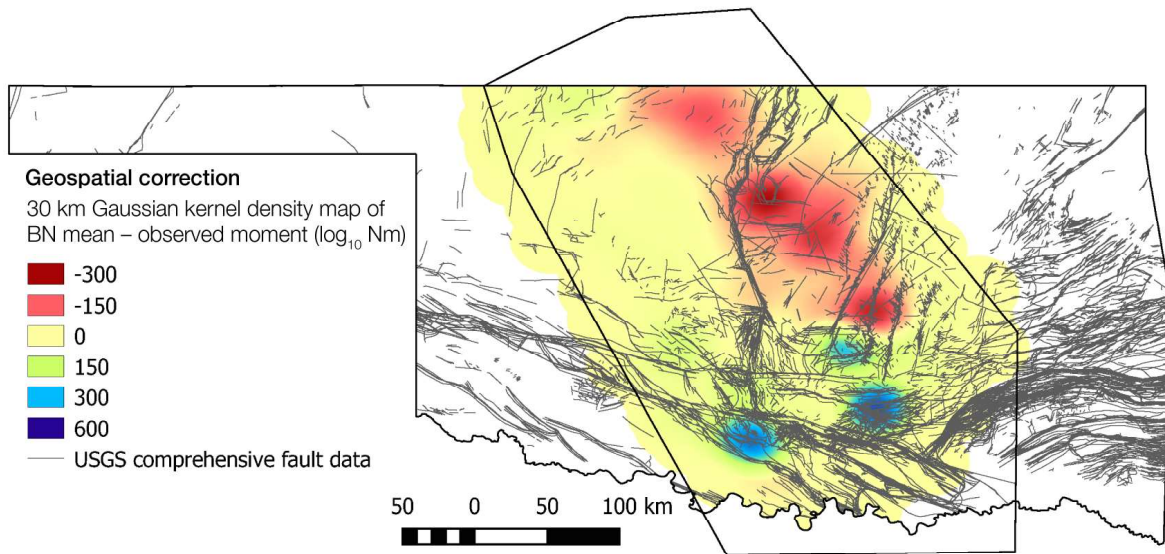


Figure S8: **Geospatial correction—mapping forecast error** | The geospatial correction is obtained by generating a kernel density map of forecast error, using a 30 km Gaussian kernel. We define the error as simply the difference between the BN mean estimate and the observed moment release (in units of \log_{10} Nm). This map shows the geospatial correction obtained from the first stage of forecasting using the unsaturated BN (Fig. 2B) and the Test 2 dataset (see Fig. S11). Note that red corresponds to areas where the first stage model initially under-forecasts and blue to areas where it over-forecasts moment release, e.g. due to geologic features or conditions that the network does not account for. Mapped faults are also shown (42); note that larger fault systems (e.g. Nemaha, Fig. 1C) appear to compartmentalize areas of under-forecasting.

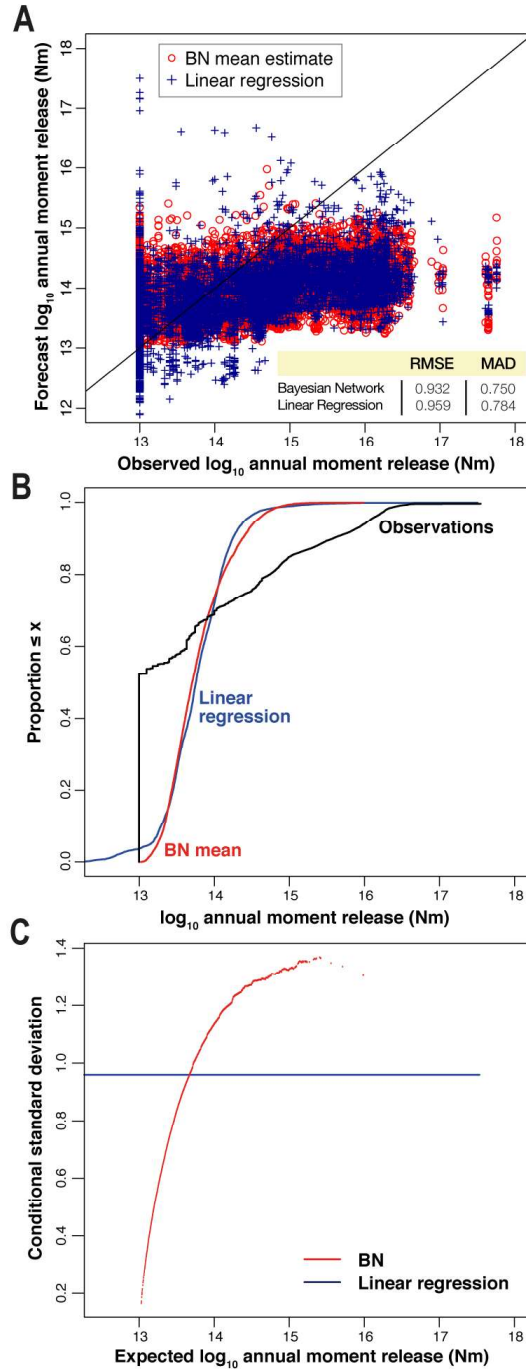


Figure S9: **Test 2 with no geospatial correction variable** | (A) Scatter plot showing forecast versus observed annual moment release for the in-sample test, using 16,000 cases drawn at random from the Test 2 dataset (see Table S3). No geospatial correction is used. Inset table shows the RMSE and MAD values for the BN and linear regression. (B) CDFs for the saturated BN and linear regression. (C) Conditional standard deviation of expected annual moment release for the BN and linear regression.

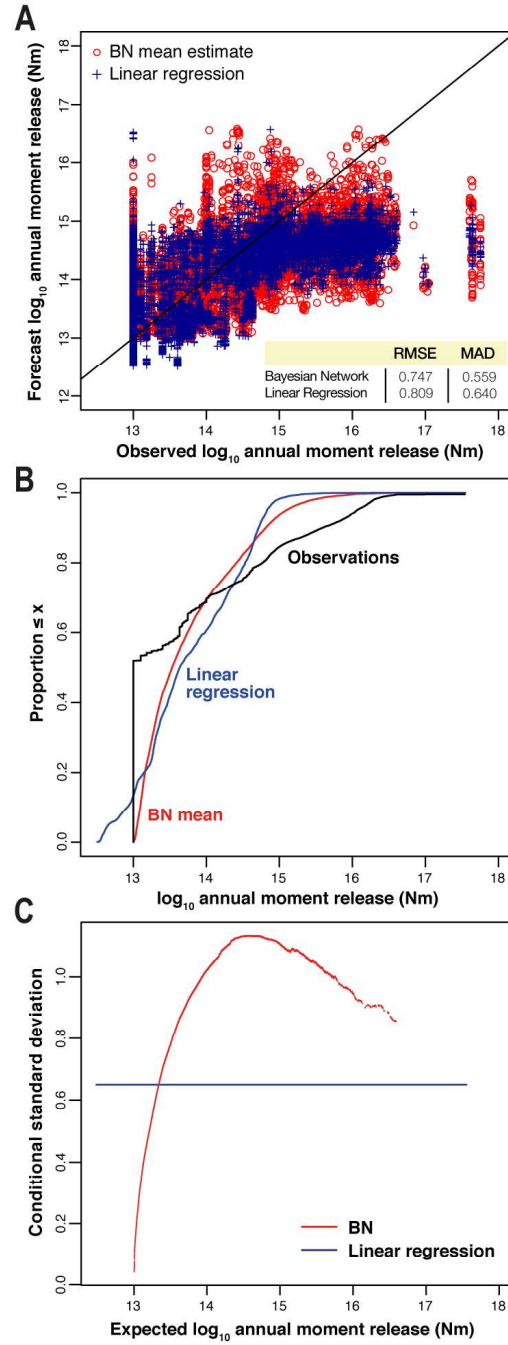


Figure S10: **Test 2 with geospatial correction variable** | (A) Scatter plot showing forecast versus observed annual moment release for the BN and regression model (both with geospatial correction). This is an in-sample test, using 16,000 cases drawn at random from the Test 2 dataset (see Table S3). Inset table shows the RMSE and MAD values for the BN and linear regression. (B) CDFs for the saturated BN and linear regression. (C) Conditional standard deviation of expected annual moment release for the BN and linear regression.

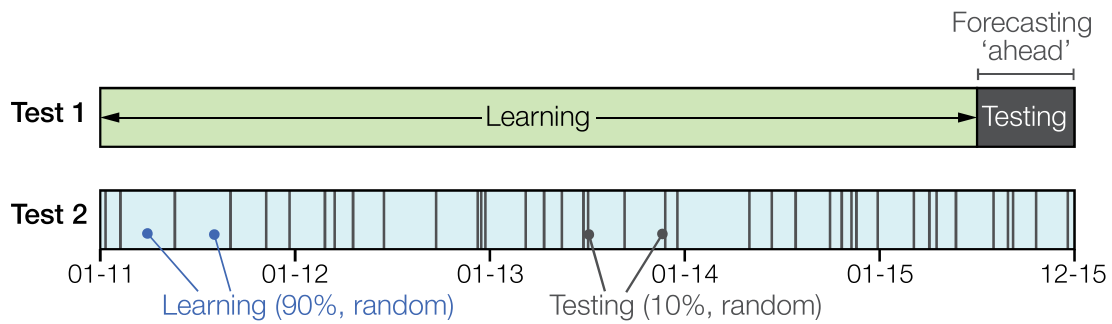


Figure S11: **Schematic illustration of the two datasets used for forecast verification** | **Test 1** uses the complete set of well records for the Oklahoma induced seismicity zone from January 2011 to June 2015 to perform the learning step, and forecasts moment for July–December 2015. Forecast verification is then carried out to evaluate how well the models perform forecasting ahead. In an operational situation, the BN could be used in this way to forecast future seismic hazard, using current or projected injection rates and depths for a specific well or region. **Test 2** uses 90% of the records from the full dataset (January 2011 to December 2015) sampled at random for learning, and the remaining 10% for forecast verification. This is used to determine how well the BN (or regression) can model seismic moment release over the past 5 years.

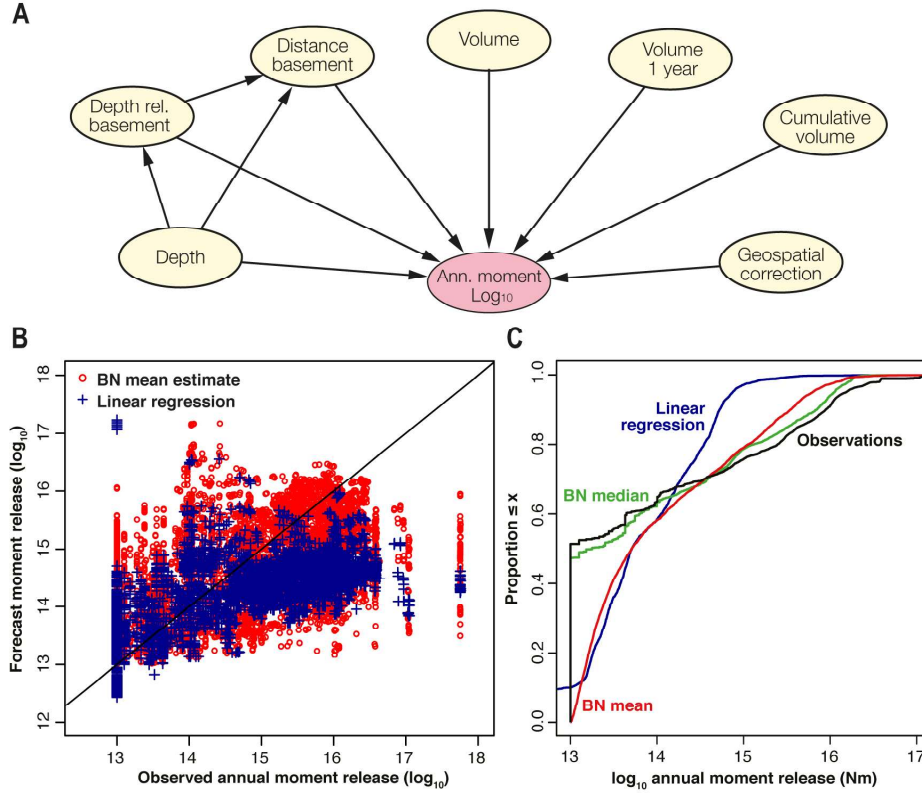
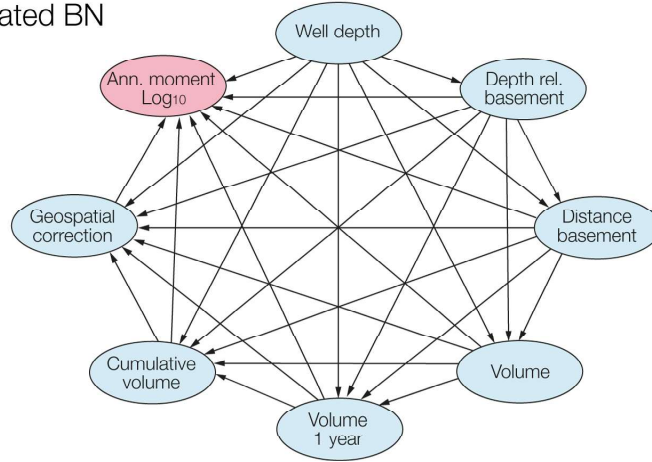
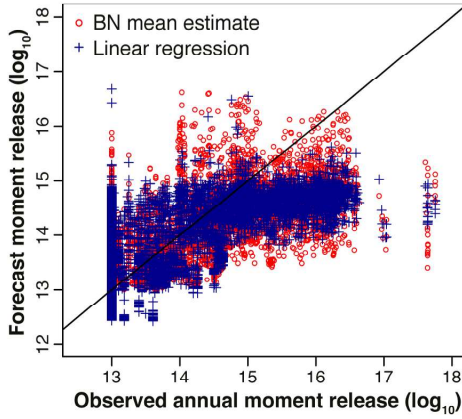


Figure S12: **Scatter and CDF plots comparing the performance of the unsaturated Bayesian Network and linear regression (Test 1)** | (A) Unsaturated BN structure. Injection rates declined in 2015, resulting in changes in relationships between cumulative, annual and monthly volume. By removing links that are no longer valid (in this case, due to the changing injection regime) the unsaturated network performs better when forecasting ahead under non-stationary conditions. (B) Scatter plot showing forecast versus observed moment release for the period June to December 2015, for the unsaturated BN and regression model. Both models under-forecast cases of very high moment release, although the BN performance is clearly better. RMSE, MAE and MAD are provided in Table S6. (C) CDFs of \log_{10} annual moment release for the BN (mean and median estimates), linear regression, and empirical data. Empirical and unconditional rank correlation matrices and regression coefficients are provided in Table S5 (see Fig. 2 and Fig. S13 for Test 2 results).

A Saturated BN



B



C

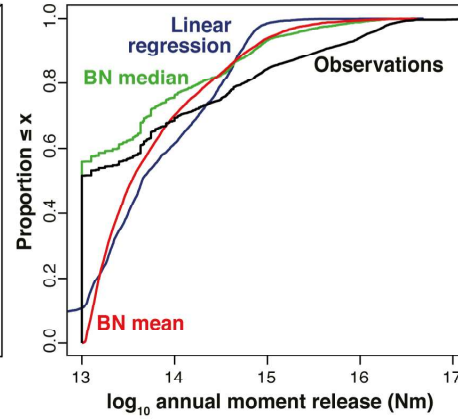


Figure S13: Scatter and CDF plots comparing the performance of the saturated Bayesian Network and linear regression (Test 2) | (A) Saturated BN structure. The saturated BN models all dependencies between variables, and performs best under stationary conditions. (B) Scatter plot showing forecast versus observed moment release for Test 2 (see Table S3 for description of the learning and testing datasets), for the unsaturated BN and regression model. Again, both models under-forecast cases of very high moment release, although the BN performance is better. RMSE, MAE and MAD are provided in Table S6. (C) CDF for the saturated network (Test 2). Empirical and unconditional rank correlation matrices and regression coefficients are provided in Table S5.

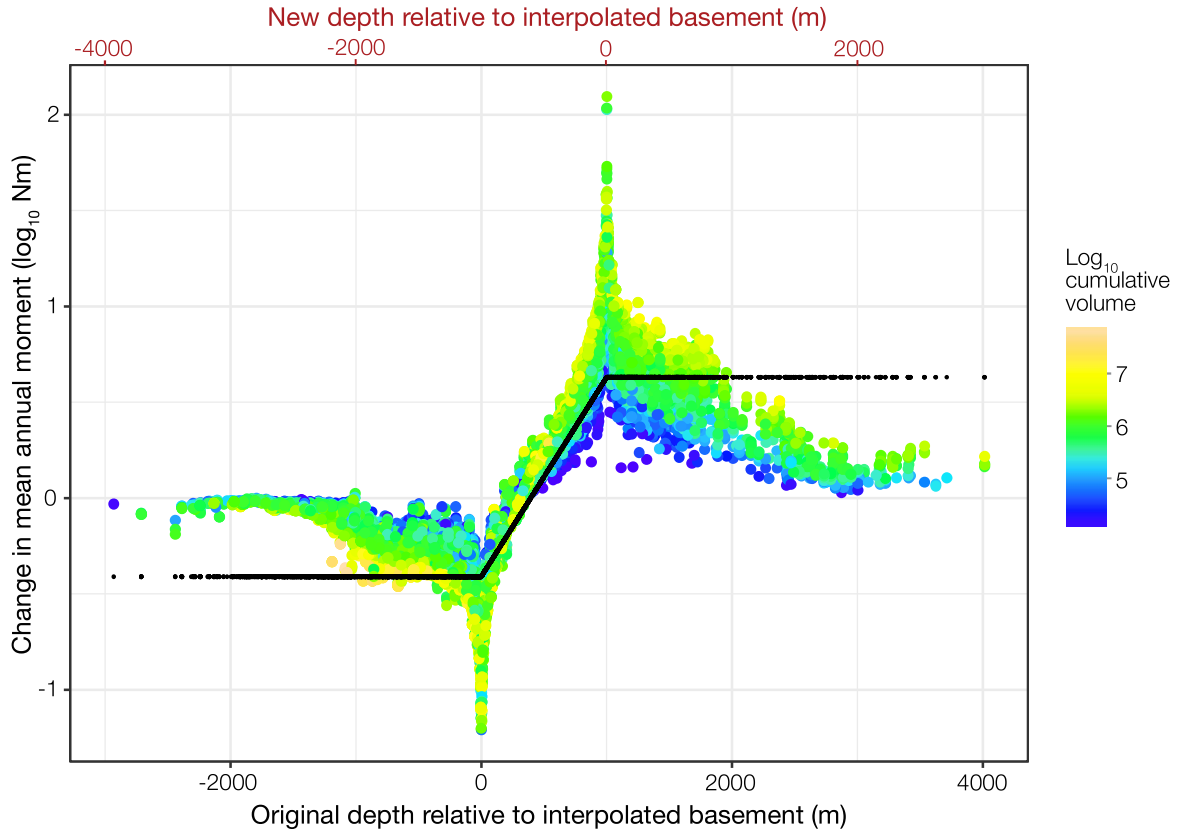


Figure S14: **Simulated effect of increasing injection well depth** | Taking a random sample of well records (10% of the full dataset), well depth is increased by 1000 m (i.e. deeper). This modified input is used to generate estimates for annual moment using the BN (in this case, using the simple unsaturated network with no spatial correction node) and linear regression models (black points). The plot shows the change in annual moment versus original well depth (see Fig. 4A for further details). The scatter (here and in Fig. 4) is due to the fact that the BN captures conditional dependency—although only the depth variables are changed, the effect of injection volume on moment release is conditional on depth. The linear model predicts the same change in moment regardless of volume (assuming volume is not changed).

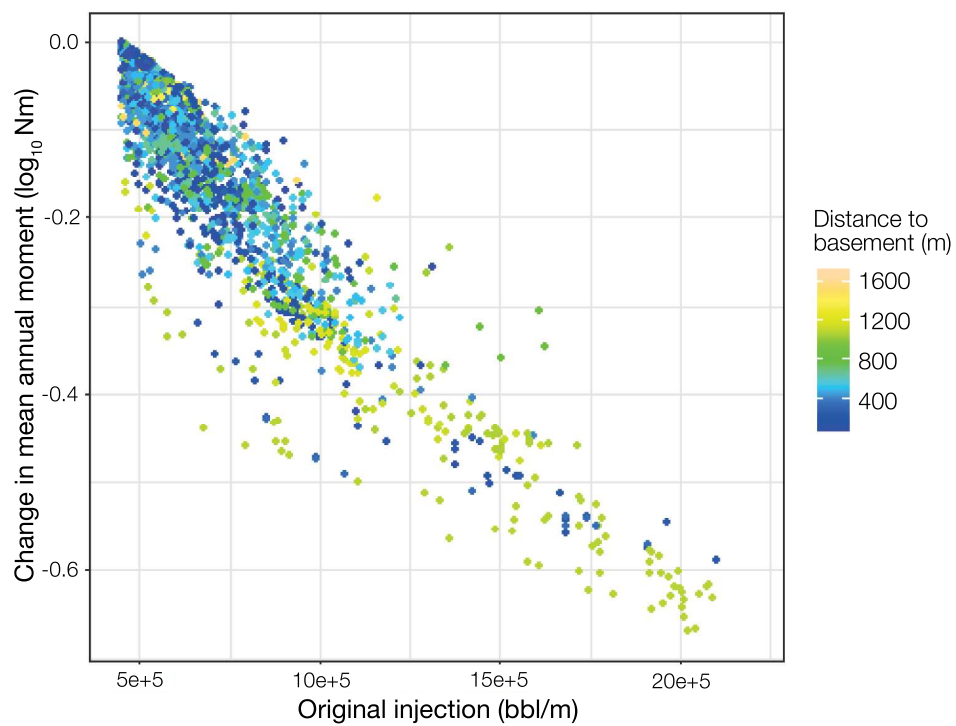


Figure S15: **Predicted change in annual seismic moment when injection volume is limited**
 | In this simulation, injection is limited to 450,000 bbl/month, and symbols are colored by distance (in meters) between the well termination and the basement interface. More explanation is provided in Fig. 4B.

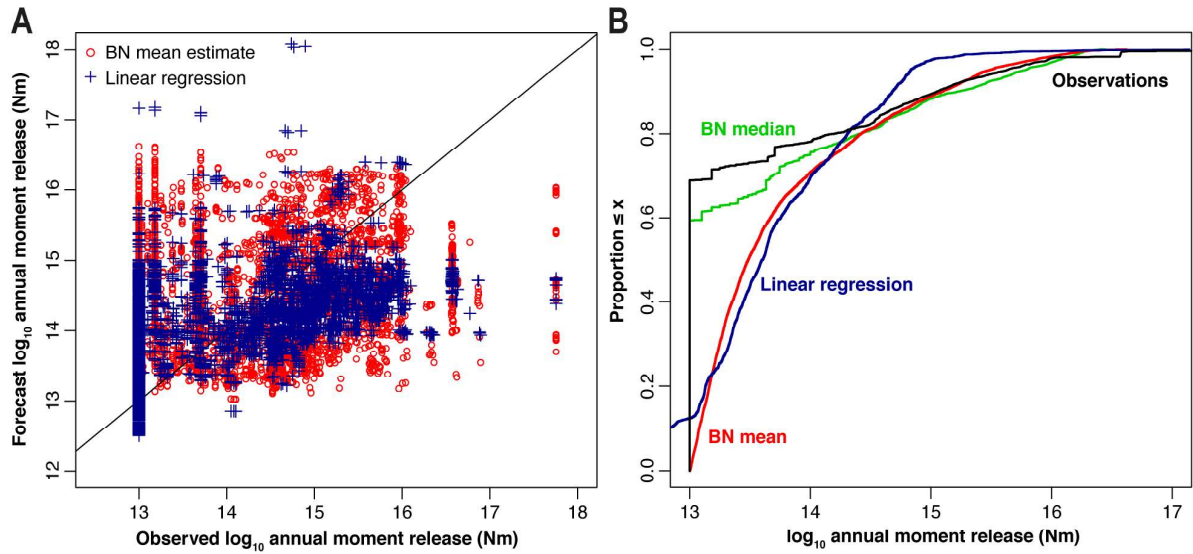


Figure S16: **Forecast verification (unsaturated BN, original geospatial correction) using injection data for 2016** | (A) Scatter plot showing forecast versus observed annual moment release for January to September 2016 for both BN and regression models. The BN forecast was generated using the unsaturated BN (structure shown in Fig. S12), learning with data from January 2011 to December 2015 and original geospatial correction (Fig. S8). (B) CDFs of \log_{10} annual moment release for the BN (mean and median estimates), linear regression, and empirical data. BN (mean) RMSE: 0.86; Linear regression RMSE: 0.79.

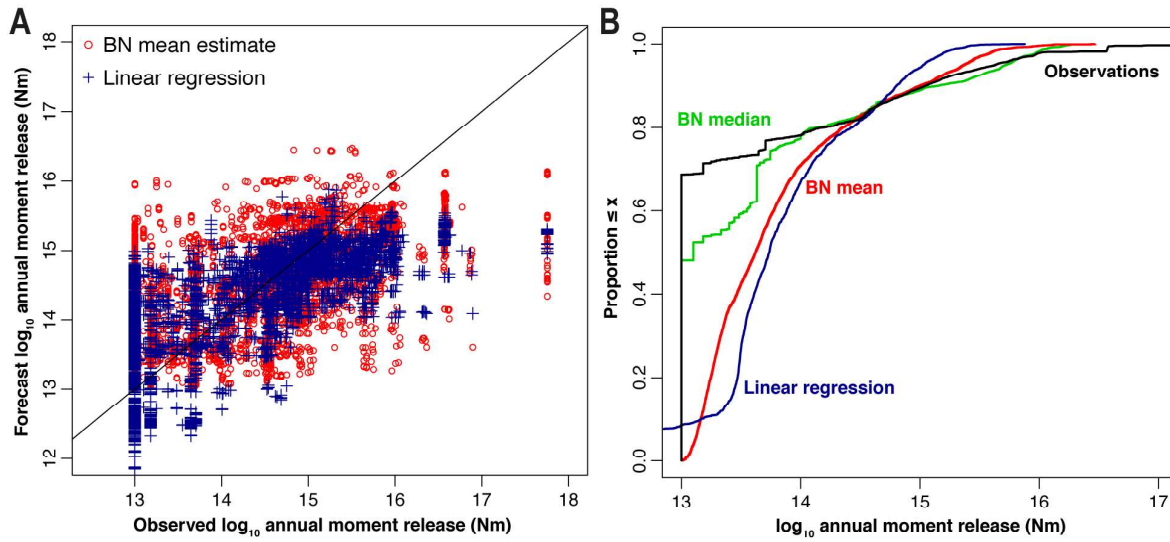


Figure S17: **Forecast verification (updated BN and geospatial correction) using injection data for 2016** | (A) Scatter plot showing forecast versus observed annual moment release for January to September 2016 for both BN and regression models. The BN forecast was generated using the unsaturated BN (structure shown in Fig. S12), learning with data from January 2014 to December 2015 and an updated geospatial correction. (B) CDFs of \log_{10} annual moment release for the BN (mean and median estimates), linear regression, and empirical data. BN (mean) RMSE: 0.73; Linear regression RMSE: 0.76.

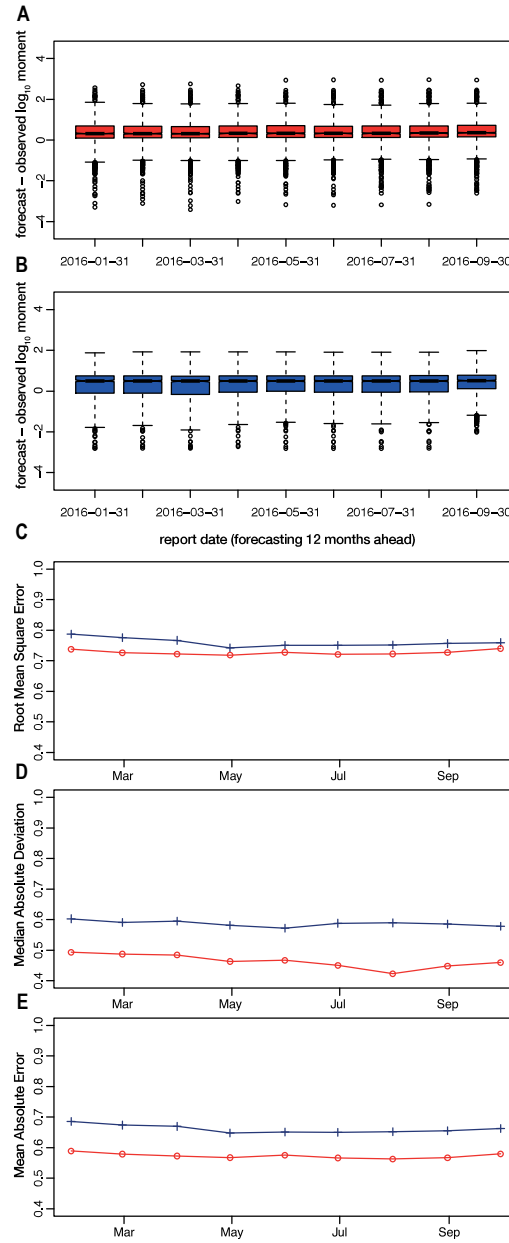


Figure S18: **Change in BN and regression forecast error for the 2016 test dataset** | (A) BN mean forecast residuals, and (B) regression model forecast residuals; learning using monthly records from January 2014 to December 2015 and forecasting annual moment release in the year ahead from January to September 2016. Both models use an updated geospatial correction (based on the 2015 forecast). The box hinges show the first and third quartiles, the middle line denotes the median, whiskers extend to the point no more than ± 1.5 times the interquartile range from the box, and the circles denote outliers beyond the whiskers. (C) RMSE BN (mean, red line) and regression (blue line) forecast. (D) MAD BN (mean) and regression forecast. (E) MAE BN (mean) and regression forecast.

Supplementary Tables

Table S1: **Injection well data** | Summary of injection well data (35) for the State of Oklahoma and the Oklahoma induced seismicity zone (Fig. 1), showing the total annual injected volume (bbl/year) and the total number of active wells (where injected volume > 0 bbl). †Please note that injection data for 2016 are preliminary.

Year Total injected vol. (bbl/yr) Number of active wells

State of Oklahoma

2011	1.84E+09	7923
2012	2.14E+09	8576
2013	2.40E+09	8793
2014	2.64E+09	8755
2015	2.61E+09	8925
2016	†2.09E+09	†7338

Oklahoma induced seismicity zone (USGS)

2011	1.44E+09	5552
2012	1.76E+09	6368
2013	2.02E+09	6543
2014	2.27E+09	6618
2015	2.24E+09	6658
2016	†1.86E+09	†5811

Number: 1	Author: vrobin03	Subject: Highlight	Date: 2/5/2020 1:29:45 PM -07'00'
Table S1: Injection well data			

Table S2: **Nodes used in the BN analysis** | Note that total moment release is calculated for $M_w > 2$ earthquakes ≤ 10 km deep.

BN node	BN node description
<i>Volume</i>	Monthly injection volume (bbl)
<i>Depth</i>	Total well depth (m)
<i>lmom_1y</i>	Total moment release (\log_{10} Nm) within 20 km of the well, within 1 year ahead
<i>Depth_rel_InterpB</i>	Well depth (m) relative to estimated basement interface: +ve (above) or -ve (below)
<i>CumVolCalc</i>	Well cumulative volume (bbl)
<i>Vsum_1y</i>	Total injected volume over the past year only, for a particular well (bbl/year)
<i>distBasement</i>	Absolute distance from the basement interface (m)
<i>SATdiffmean</i>	Geospatial correction node for the saturated model
<i>UNSATdiffmean</i>	Geospatial correction node for the unsaturated model
<i>LRdiffmean</i>	Geospatial correction node for the linear regression model

Table S3: **Number of cases for learning and training** | Table shows the total number of individual monthly records for all wells (i.e. a total of 2,897) used in the BN analysis. Test 1—learning with records from January 2011 to June 2015, forecasting July to December 2015. Test 2—learning with a random sample of 90% of records from January 2011 to December 2015; forecasting with the remaining 10%. In-sample test—uses the Test 2 dataset (95,673 case) for learning, and a random sample of 16,000 cases drawn from the learning set to visualize results (see Fig. S11). Cases with $lmom_1y > 16.5$ are in the tail end of the distribution used for learning the network. Forecasting ahead (Test 1) the network learns with just 0.5% of cases with $lmom_1y > 16.5$; but the forecast verification data (July to December 2015) has 2% of cases with $lmom_1y > 16.5$.

	Total cases	lmom_1y>16		lmom_1y>16.5	
Test 1 learning	94,905	5,235	5.50%	513	0.50%
Test 1 forecasting	11,399	1,011	8.90%	225	2.00%
Test 2 learning	95,673	5,635	5.90%	672	0.70%
Test 2 forecasting	10,631	611	5.70%	66	0.60%
In-sample testing	16,000	965	6.0%	116	0.7%

Table S4: **Effect of node hierarchy** | Investigation of the effect of node hierarchy on conditional rank correlation for the distance to basement and \log_{10} moment nodes (this example uses the saturated BN, Test 2 data, 90% of full dataset). Given all other covariates, the conditional rank correlation for distance to basement and \log_{10} moment remains the highest, and changes very little under the alternate ordering (see Table S2 for definitions of variables).

Conditional Rank Correlation Coefficients	Value
Depth \log_{10} moment	0.16587
Depth _{rel} InterpB \log_{10} moment Depth	-0.13467
distBasement \log_{10} moment Depth Depth_{rel}InterpB	-0.29804
Volume \log_{10} moment Depth Depth _{rel} InterpB distBasement	0.11851
Vsum \log_{10} moment Depth Depth _{rel} InterpB distBasement Volume	0.11246
CumVol \log_{10} moment Depth Depth _{rel} InterpB distBasement Volume Vsum	0.16238
Conditional Rank Correlation Coefficients—alternate hierarchy	Value
Depth \log_{10} moment	0.16587
Depth _{rel} InterpB \log_{10} moment Depth	-0.13467
Volume \log_{10} moment Depth Depth _{rel} InterpB	0.16293
Vsum \log_{10} moment Depth Depth _{rel} InterpB Volume	0.11764
CumVol \log_{10} moment Depth Depth _{rel} InterpB Volume Vsum	0.14722
distBasement \log_{10} moment Depth Depth_{rel}InterpB Volume Vsum CumVol	-0.28354

Table S5: **Unconditional rank correlation matrices and regression coefficients** | Empirical rank correlation (unconditional) matrix, Bayesian Network unconditional rank correlation matrices (for the saturated and unsaturated networks), and linear regression coefficients.

Empirical Rank Correlation Matrix (Test 2, calculated using random 90% of full dataset from 2011–15)									
	Depth	Depth rel. basement	Distance basement	Volume	Volume 1 year	Cumulative volume	Geospatial correction	Annual moment	\log_{10}
Depth	1	-0.18	-0.20	+0.23	+0.19	+0.14	-0.28	+0.20	
Depth rel. basement		1	+0.29	-0.02	-0.02	-0.02	-0.22	-0.20	
Distance basement			1	-0.20	-0.18	-0.12	+0.44	-0.34	
Volume				1	+0.82	+0.63	-0.22	+0.18	
Volume 1 year					1	+0.85	-0.19	+0.20	
Cumulative volume						1	-0.14	+0.26	
Geospatial correction							1	-0.55	
Annual moment \log_{10}								1	
BN Rank Correlation Matrix (saturated network, Test 2, calculated using random 90% of full dataset from 2011–15)									
	Depth	Depth rel. basement	Distance basement	Volume	Volume 1 year	Cumulative volume	Geospatial correction	Annual moment	\log_{10}
Depth	1	-0.17	-0.16	+0.22	+0.18	+0.14	-0.21	+0.17	
Depth rel. basement		1	+0.25	-0.03	-0.03	-0.02	-0.26	-0.16	
Distance basement			1	-0.19	-0.17	-0.13	+0.37	-0.34	
Volume				1	+0.79	+0.62	-0.21	+0.19	
Volume 1 year					1	+0.87	-0.18	+0.22	
Cumulative volume						1	-0.14	+0.26	
Geospatial correction							1	-0.52	
Annual moment \log_{10}								1	
BN Rank Correlation Matrix (unsaturated network, Test 1, calculated using data from Jan 2011–June 2015)									
	Depth	Depth rel. basement	Distance basement	Volume	Volume 1 year	Cumulative volume	Geospatial correction	Annual moment	\log_{10}
Depth	1	-0.16	-0.16	0	0	0	0	+0.15	
Depth rel. basement		1	+0.25	0	0	0	0	-0.17	
Distance basement			1	0	0	0	0	-0.32	
Volume				1	0	0	0	+0.11	
Volume 1 year					1	0	0	+0.10	
Cumulative volume						1	0	+0.18	
Geospatial correction							1	-0.47	
Annual moment \log_{10}								1	
Regression coefficients									
	Intercept	Depth	Depth rel. basement	Distance basement	Volume	Volume 1 year	Cumulative volume	Geospatial correction	
Linear regression with spatial node	+1.46E+01	-1.90E-04	-2.81E-04	-1.23E-04	-1.60E-07	+7.59E-08	+1.45E-08	-2.89E-03	
Linear regression, no spatial node	+1.39E+01	+1.02E-04	-8.80E-06	-5.20E-04	+2.21E-07	+1.09E-07	+1.11E-08	—	

Table S6: **Errors for the BN and regression models** | Root mean squared error (RMSE), mean absolute error (MAE), and median average deviation (MAD) for the BN and regression models using both Test 1 (forecasting ‘ahead’ for July to December 2015) and Test 2 datasets. This shows the improvement in both the BN and regression model with the addition of the geospatial correction node. It can be seen that the saturated BN provides the best representation of the whole dataset, but that the unsaturated BN works better when forecasting ‘out-of-sample’, under a changed injection regime.

Test 1—BN and regression model learning with observations from Jan–June 2015, testing July–Dec 2015			
	RMSE	MAE	MAD
Linear regression – no geospatial variable	1.049	0.850	0.716
BN mean – no geospatial variable (saturated net)	1.071	0.910	0.836
BN median – no geospatial variable (saturated net)	1.076	0.781	0.633
BN mean – no geospatial variable (unsaturated net)	1.022	0.827	0.695
BN median – no geospatial variable (unsaturated net)	1.054	0.697	0.519
<hr/>			
Linear regression – with geospatial variable	0.871	0.705	0.578
BN mean estimate (saturated net) – with geospatial variable	0.914	0.737	0.642
BN median estimate (saturated net) – with geospatial variable	0.894	0.613	0.519
BN mean estimate (unsaturated net) – with geospatial variable	0.842	0.615	0.424
BN median estimate (unsaturated net) – with geospatial variable	0.861	0.526	0.234
<hr/>			
Test 2—learning with 90% of dataset, testing with 10% (randomly selected from Jan 2011–Dec 2015)			
	RMSE	MAE	MAD
Linear regression – no geospatial variable	0.947	0.778	0.698
BN mean – no geospatial variable (saturated net)	0.922	0.738	0.597
BN median – no geospatial variable (saturated net)	1.009	0.649	0.401
<hr/>			
Linear regression – with geospatial variable	0.786	0.632	0.520
BN mean estimate (saturated net) – with geospatial variable	0.733	0.554	0.425
BN median estimate (saturated net) – with geospatial variable	0.756	0.475	0.232
BN mean estimate (unsaturated net) – with geospatial variable	0.797	0.578	0.423
BN median estimate (unsaturated net) – with geospatial variable	0.842	0.530	0.295

Table S7: **Empirical and BN unconditional rank correlations for $lmom_{1y}$ calculated using preliminary injection data from January 2011 to September 2016** | Compare with correlations for January 2011 to December 2015 shown in Table S5.

	Empirical	Unsaturated BN	Saturated BN
Depth	0.23	0.19	0.19
Depth rel. basement	-0.17	-0.13	-0.13
Distance to basement	-0.35	-0.34	-0.34
Volume	0.18	0.1	0.19
Volume 1 year	0.2	0.1	0.22
Cumulative volume	0.23	0.11	0.24
Geospatial correction	-0.59	-0.46	0.57

References and Notes

1. C. Langenbruch, M. D. Zoback, How will induced seismicity in Oklahoma respond to decreased saltwater injection rates? *Sci. Adv.* **2**, e1601542 (2016). [Medline](#)
2. W. L. Ellsworth, Injection-induced earthquakes. *Science* **341**, 1225942 (2013). [Medline](#)
3. K. M. Keranen, M. Weingarten, G. A. Abers, B. A. Bekins, S. Ge, Induced earthquakes. Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. *Science* **345**, 448–451 (2014). [doi:10.1126/science.1255802](#) [Medline](#)
4. M. Weingarten, S. Ge, J. W. Godt, B. A. Bekins, J. L. Rubinstein, High-rate injection is associated with the increase in U.S. mid-continent seismicity. *Science* **348**, 1336–1340 (2015). [doi:10.1126/science.aab1345](#) [Medline](#)
5. M. D. Petersen *et al.*, Incorporating induced seismicity in the 2014 United States national seismic hazard model—results of 2014 workshop and sensitivity studies, Tech. Rep. 20151070, U.S. Geological Survey (2015).
6. M. D. Petersen *et al.*, 2016 one-year seismic hazard forecast for the Central and Eastern United States from induced and natural earthquakes, Tech. Rep., U.S. Geological Survey, Reston, VA (2016).
7. F. R. Walsh 3rd, M. D. Zoback, Oklahoma’s recent earthquakes and saltwater disposal. *Sci. Adv.* **1**, e1500195 (2015). [Medline](#)
8. P. O. Ogwari, S. P. Horton, Numerical model of pore-pressure diffusion associated with the initiation of the 2010–2011 Guy–Greenbrier, Arkansas earthquakes. *Geofluids* **16**, 954–970 (2016). [doi:10.1111/gfl.12198](#)
9. A. K. Shah, G. R. Keller, Geologic influence on induced seismicity: Constraints from potential field data in Oklahoma. *Geophys. Res. Lett.* **44**, 152–161 (2017). [doi:10.1002/2016GL071808](#)
10. N. D. McMahon, R. C. Aster, W. L. Yeck, D. E. McNamara, H. M. Benz, Spatio-temporal evolution of the 2011 Prague, Oklahoma aftershock sequence revealed using subspace detection and relocation. *Geophys. Res. Lett.* **44**, 7149–7158 (2017). [doi:10.1002/2017GL072944](#)
11. K. M. Keranen, H. M. Savage, G. A. Abers, E. S. Cochran, Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology* **41**, 699–702 (2013). [doi:10.1130/G34045.1](#)
12. W.-Y. Kim, Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio. *J. Geophys. Res. Solid Earth* **118**, 3506–3518 (2013). [doi:10.1002/jgrb.50247](#)
13. S. Horton, Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in Central Arkansas with potential for damaging earthquake. *Seismol. Res. Lett.* **83**, 250–260 (2012). [doi:10.1785/gssrl.83.2.250](#)
14. D. E. McNamara, H. M. Benz, R. B. Herrmann, E. A. Bergman, P. Earle, A. Holland, R. Baldwin, A. Gassner, Earthquake hypocenters and focal mechanisms in central

- Oklahoma reveal a complex system of reactivated subsurface strike-slip faulting. *Geophys. Res. Lett.* **42**, 2742–2749 (2015). [doi:10.1002/2014GL062730](https://doi.org/10.1002/2014GL062730)
15. M. Shirzaei, W. L. Ellsworth, K. F. Tiampo, P. J. González, M. Manga, Surface uplift and time-dependent seismic hazard due to fluid injection in eastern Texas. *Science* **353**, 1416–1419 (2016). [doi:10.1126/science.aag0262](https://doi.org/10.1126/science.aag0262) [Medline](#)
 16. E. J. Fielding, S. S. Sangha, D. P. S. Bekaert, S. V. Samsonov, J. C. Chang, Surface deformation of North-Central Oklahoma related to the 2016 Mw 5.8 Pawnee earthquake from SAR interferometry time series. *Seismol. Res. Lett.* **88**, 971–982 (2017). [doi:10.1785/0220170010](https://doi.org/10.1785/0220170010)
 17. M. Schoenball, W. L. Ellsworth, Waveform-relocated earthquake catalog for Oklahoma and Southern Kansas illuminates the regional fault network. *Seismol. Res. Lett.* **88**, 1252–1258 (2017). [doi:10.1785/0220170083](https://doi.org/10.1785/0220170083)
 18. X. Chen, N. Nakata, C. Pennington, J. Haffener, J. C. Chang, X. He, Z. Zhan, S. Ni, J. I. Walter, The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Sci. Rep.* **7**, 4945 (2017). [doi:10.1038/s41598-017-04992-z](https://doi.org/10.1038/s41598-017-04992-z) [Medline](#)
 19. National Research Council (U.S.), *Induced Seismicity Potential in Energy Technologies* (National Academy of Sciences, Washington, DC, 20001, 2013).
 20. Oklahoma Corporation Commission Earthquake Response Summary, 18 Nov 2016; www.occeweb.com/News/2016/11-23-16EARTHQUAKE%20ACTION%20SUMMARY.pdf (2016).
 21. M. D. Petersen, C. S. Mueller, M. P. Moschetti, S. M. Hoover, A. M. Shumway, D. E. McNamara, R. A. Williams, A. L. Llenos, W. L. Ellsworth, A. J. Michael, J. L. Rubinstein, A. F. McGarr, K. S. Rukstales, 2017 one-year seismic-hazard forecast for the Central and Eastern United States from induced and natural earthquakes. *Seismol. Res. Lett.* **88**, 772–783 (2017). [doi:10.1785/0220170005](https://doi.org/10.1785/0220170005)
 22. *Sierra Club v. Chesapeake Operating LLC et al.*, F. Supp. 3d, WL 1287546 (W.D. Okla. 2017).
 23. D. E. McNamara, G. P. Hayes, H. M. Benz, R. A. Williams, N. D. McMahon, R. C. Aster, A. Holland, T. Sickbert, R. Herrmann, R. Briggs, G. Smoczyk, E. Bergman, P. Earle, Reactivated faulting near Cushing, Oklahoma: Increased potential for a triggered earthquake in an area of United States strategic infrastructure. *Geophys. Res. Lett.* **42**, 8328–8332 (2015). [doi:10.1002/2015GL064669](https://doi.org/10.1002/2015GL064669)
 24. S. Barrett, The link between hydrofracking, wastewater injection and earthquakes: key issues for re/insurers, Tech. Rep., Swiss Reinsurance (2016). www.swissre.com/library/2016_08_link_hydrofracking_wastewater_injection_earthquakes_key_issues_reinsurers.html.
 25. Oklahoma Corporation Commission, Central Oklahoma Regional Volume Reduction Plan and Expansion of Area of Interest; www.occeweb.com/News/2016/03-07-16ADVISORY-AOI,%20VOLUME%20REDUCTION.pdf (2016).

26. M. van der Baan, F. J. Calixto, Human-induced seismicity and large-scale hydrocarbon production in the USA and Canada. *Geochem. Geophys. Geosyst.* **18**, 2467–2485 (2017). [doi:10.1002/2017GC006915](https://doi.org/10.1002/2017GC006915)
27. T. H. W. Goebel, J. I. Walter, K. Murray, E. E. Brodsky, Comment on “How will induced seismicity in Oklahoma respond to decreased saltwater injection rates?” by C. Langenbruch and M. D. Zoback. *Sci. Adv.* **3**, e1700441 (2017). [doi:10.1126/sciadv.1700441](https://doi.org/10.1126/sciadv.1700441) [Medline](#)
28. A. J. Barbour, J. H. Norbeck, J. L. Rubinstein, The effects of varying injection rates in Osage County, Oklahoma, on the 2016 Mw 5.8 Pawnee earthquake. *Seismol. Res. Lett.* **88**, 1040–1053 (2017). [doi:10.1785/0220170003](https://doi.org/10.1785/0220170003)
29. Oklahoma Corporation Commission directive, <http://occeweb.com/News/DIRECTIVE-2.pdf> (2015).
30. Y. Zhang, M. Person, J. Rupp, K. Ellett, M. A. Celia, C. W. Gable, B. Bowen, J. Evans, K. Bandilla, P. Mozley, T. Dewers, T. Elliot, Hydrogeologic controls on induced seismicity in crystalline basement rocks due to fluid injection into basal reservoirs. *Ground Water* **51**, 525–538 (2013). [doi:10.1111/gwat.12071](https://doi.org/10.1111/gwat.12071) [Medline](#)
31. B. S. Currie, M. Brudzinski, R. Skoumal, Geological Society of America Abstracts with Programs (2016), vol. 48.
32. Materials and methods are available as supplementary materials.
33. UNINET software designed by the Risk and Environmental Modeling group, Delft University of Technology, developed by Dan Ababei, Lighttwist Software; www.lighttwist.net/wp/uninet.
34. The R Project for Statistical Computing, <https://cran.r-project.org/>.
35. Oklahoma Corporation Commission, UIC well data from www.occeweb.com/og/ogdatafiles2.htm.
36. ANSS Comprehensive Earthquake Catalog (ComCat), U.S. Geological Survey, <https://earthquake.usgs.gov/data/comcat/>; accessed 8 March 2017.
37. A. McGarr, A. J. Barbour, Wastewater disposal and the earthquake sequences during 2016 near Fairview, Pawnee, and Cushing, Oklahoma. *Geophys. Res. Lett.* **44**, 9330–9336 (2017). [doi:10.1002/2017GL075258](https://doi.org/10.1002/2017GL075258)
38. QGIS Development Team, QGIS Geographic Information System, <https://qgis.org> (2016).
39. J. Campbell, J. Weber, Wells drilled to basement in Oklahoma, Tech. Rep. 2006-1, Oklahoma Geological Survey, <http://ogs.ou.edu/docs/specialpublications/SP2006-1.pdf> (2006).
40. E. K. Franseen, A review of Arbuckle Group strata in Kansas from a sedimentologic perspective: Insights for future research from past and recent studies, Kansas Geological Survey, Open-file Report 1999-49 75, 68 (2000).
41. R. A. Northcutt, J. A. Campbell, in *Geologic Provinces of Oklahoma* (Springer Netherlands, Dordrecht, 1998), pp. 29–37.

42. S. Marsh, A. Holland, Comprehensive fault database and interpretive fault map of Oklahoma, Tech. Rep. OF2-2016, Oklahoma Geological Survey, The University of Oklahoma, Norman, OK 73019 (2016).
43. Kansas Corporation Commission, Second Order Reducing Saltwater Injection Rates; www.kcc.state.ks.us/pi/press/16-11.htm (2016).
44. W. L. Yeck, G. P. Hayes, D. E. McNamara, J. L. Rubinstein, W. D. Barnhart, P. S. Earle, H. M. Benz, Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts. *Geophys. Res. Lett.* **44**, 711–717 (2017). [doi:10.1002/2016GL071685](https://doi.org/10.1002/2016GL071685)
45. W. L. Yeck, M. Weingarten, H. M. Benz, D. E. McNamara, E. A. Bergman, R. B. Herrmann, J. L. Rubinstein, P. S. Earle, Far-field pressurization likely caused one of the largest injection induced earthquakes by reactivating a large preexisting basement fault structure. *Geophys. Res. Lett.* **43**, 10,198–10,207 (2016). [doi:10.1002/2016GL070861](https://doi.org/10.1002/2016GL070861)
46. A. M. Hanea, D. Kurowicka, R. M. Cooke, Hybrid method for quantifying and analyzing Bayesian belief nets. *Qual. Reliab. Eng. Int.* **22**, 709–729 (2006). [doi:10.1002/qre.808](https://doi.org/10.1002/qre.808)
47. D. Kurowicka, R. Cooke, in *Modern Statistical and Mathematical Methods in Reliability*, A. Wilson, N. Limnios, S. Keller-McNulty, Y. Armijo, Eds. (World Scientific, 2005), pp. 309–323.
48. S. Wiemer, M. Wyss, Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bull. Seismol. Soc. Am.* **90**, 859–869 (2000). [doi:10.1785/0119990114](https://doi.org/10.1785/0119990114)
49. Rcomcat library for searching ANSS ComCat comprehensive earthquake catalog; <https://github.com/usgs/rcomcat>.
50. S. D. Davis, C. Frohlich, Did (or will) fluid injection cause earthquakes? Criteria for a rational assessment. *Seismol. Res. Lett.* **64**, 207 (1993).
51. A. Frankel, Mapping seismic hazard in the Central and Eastern United States. *Seismol. Res. Lett.* **66**, 8–21 (1995). [doi:10.1785/gssrl.66.4.8](https://doi.org/10.1785/gssrl.66.4.8)
52. J. Pearl, Causality: Models, Reasoning and Inference (Cambridge Univ. Press, ed. 2, 2009).
53. The R Project for Statistical Computing VineCopula package; <https://cran.r-project.org/web/packages/VineCopula/VineCopula.pdf>.
54. R. M. Cooke, H. Joe, B. Chang, *Vine Regression*, Resources for the Future Discussion Paper (2015), pp. 15–52.